

# Случайные процессы. Прикладной поток.

## Практическое задание 8.

Прогнозирование временных рядов.



### Правила:

- Выполненную работу нужно отправить на почту [probability.diht@yandex.ru](mailto:probability.diht@yandex.ru), указав тему письма "[СП17] Фамилия Имя - Задание 8". Квадратные скобки обязательны. Вместо Фамилия Имя нужно подставить свои фамилию и имя.
- Прислать нужно ноутбук и его pdf-версию. Названия файлов должны быть такими: 8.N.ipynb и 8.N.pdf, где N — ваш номер из таблицы с оценками.
- Никакой код из данного задания при проверке запускаться не будет.
- При выполнении задания можно использовать код с семинара. Во всяком случае ноутбук точно стоит посмотреть.

В файле *electricity.csv* (отсюда) содержится информация о максимальном спросе на электричество (Consumption) в штате Виктория (Австралия) за 30-минутные интервалы с 10 января 2000 в течении 115 дней, а так же информация о температуре воздуха (Temperature) за эти же промежутки времени.

### Задание:

1. Нарисуйте графики временных рядов температуры и потребления электричества. Верно ли, что спрос на электричество зависит от температуры воздуха? Для ответа на вопрос используйте коэффициенты корреляции, учитывая условия их применимости.
2. Разделите временной ряд на две части: данные за последнюю неделю (последние 48\*7 измерений) назовем тестовыми данными, а все остальное — обучающими данными. Пункты задания 3-6 выполните для обучающих данных.

3. Сколько типов сезонностей можно выделить в каждом из двух рядов (спрос на электричество и температура)? С помощью STL-декомпозиции в каждом ряде выделите тренд, все типы сезонности, остатки.
4. С помощью критерия KPSS проверьте на стационарность исходные ряды и остатки, полученные после применения STL-декомпозиции. Не забывайте про множественную проверку гипотез.
5. С помощью преобразований исходных рядов приведите их к стационарным. По графикам ACF и PACF подберите параметры модели  $SARIMA(p, d, q) \times (P, D, Q)_s$ .
6. С помощью поиска по сетке вокруг выбранных параметров подберите оптимальные параметры по значению AIC. Учтите, что из сделанных ранее преобразований ряда нужно оставить лишь некоторые. Другие, например, одна из сезонностей будут учтены параметрами модели.
7. Постройте прогнозы модели с оптимальными параметрами на неделю вперед. Посчитайте качество прогноза по сравнению с реальными данными на тестовом интервале, используя метрику MSE (см. презентацию).
8. Добавьте в модель предсказания электричества экзогенные факторы:
  - (a) Дневную и месячную сезонность (очевидно, они известны заранее). Однако, в том виде как они записаны в таблице применять не хорошо — может работать плохо, поэтому стоит использовать гармоники Фурье — синусы с периодом, делящим период сезонности. Их использование может позволить учесть сложные сезонности.
  - (b) Значения температуры, используя на тестовом интервале времени истинные значения температуры (нечестный способ).
  - (c) Значения температуры, используя на тестовом интервале времени предсказания значений температуры.
  - (d) Вместе (a) и (c).
  - (e) \* Использование значений температуры по частям — для получения прогноза  $\hat{y}_{T+h|T}$  строится своя модель по временному ряду  $y_h, \dots, y_T$  с рядом экзогенного фактора  $x_1, \dots, x_{T-h}$ . Тогда для получения прогноза  $\hat{y}_{T+h|T}$  нужно знать значения  $x_{T-h+1}, \dots, x_T$ , которые известны на момент построения модели.
  - (f) \* Вместе (a) и (e).
9. Сравните все предсказания по метрике MSE.